#6
Linda
11/27/00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Arnold et al.	Group Art Unit: 1655
Serial No. 09/586,156	Examiner: Lu, F.
Filed: June 2, 2000	Attorney Docket No. IN-0016-1
For: *Duplex Probes for Hybridization Reactions*	

DECLARATION UNDER 37CFR1.132

I, Mark A. Reynolds, declare and state as follows:

1. I am the Senior Director of Assay Research at Incyte Genomics, Inc., the assignee of the subject patent application. I hold a Ph.D. in pharmaceutical chemistry and have over 17 years industrial experience in nucleic acid research, and am knowledgeable and experienced in the field of microarrays. I have read and am familiar with the contents of the above application and the office action mailed 9/13/00.

2. Claim 1 requires only a single step: combining a polynucleotide probe with a polynucleotide target stably associated with a surface of a solid support, wherein one of the probe and target is double-stranded comprising complementary strands, and the other is single-stranded having complementarity with one of the complementary strands, under conditions wherein the probe and target hybridize and the probe is thereby immobilized. Every aspect of this claimed method is amply described, including methods for making the probes (p.8, line 12 - p.9, line 10) and targets (p.6, line 19 - p.7, line 22); methods for stably associating the targets to a solid support surface (p.7, line 23 - p.8, line 11) and conditions for hybridizing the probe to the immobilized target (p.10, lines 5-8). In addition, the entire method is repeatedly exemplified in experimental detail (p.12, line 5 - p.18, line 10).

Claim 3 is the method of claim 1 further comprising the step of releasing the immobilized probe. This is trivial - the additional step merely requires subjecting the probe to conditions whereby it would be released from the immobilized target. Conditions for hybridizing and

dehybridizing polynucleotides are routine, well-known in the art and involve no more than adjusting temperature and/or salt concentrations (e.g. Maniatis et al., cited on p.10, line 6; see also p.17, line 11). The specification readily enables those skilled in the art to hybridize the probe to the immobilized target; with or without our specification, they are readily able to dehybridize it.

Claim 4 is the method of claim 1 further comprising the step of amplifying the immobilized probe. Methods for amplifying a given polynucleotide are routine and well-established (e.g. Maniatis et al., cited on p.8, line 22); amplification is in fact how the probes are generally made (p.8, line 28) – the targets too (as amplified PCR elements, p.17, lines 4-5).

Claim 5 is the method of claim 1 further comprising the step of amplifying the immobilized probe and then detecting resultant amplified probe. The additional step (over claim 4) of detecting the amplified probe is routine: methods for labeling amplified probes are described (p.8, lines 27-28) as are methods for detecting them (p.10, lines 8-11).

Claim 6 is the method of claim 1 further comprising the steps of releasing and amplifying the probe to produce a labeled, amplified double-stranded probe, hybridizing the labeled probe to a target polynucleotide to immobilize the labeled probe and detecting the labeled probe. The additional step (over claims 3 and 5) of hybridizing the probe to an immobilized target polynucleotide requires no more than repeating the step of claim 1. Hybridizing probes to immobilized targets is thoroughly described and exemplified (see e.g. Examples 1 and 2).

Claim 7 is the method of claim 1 further comprising the steps of releasing and amplifying the probe to produce a labeled, amplified double-stranded probe, hybridizing the labeled probe to a target polynucleotide to immobilize the labeled probe and detecting the labeled probe, wherein the probe is double-stranded and target is single-stranded. The additional limitation (over claim 7) is the requirement that the probe is double-stranded and target is single-stranded, as exemplified in Examples 3-10.

How to isolate or release an immobilized probe, how to amplify such probe and how to detect an amplified probe are all trivial tasks to those skilled in the art of solid phase hybridization assays. In my expert opinion, the teachings and exemplification of the specification fully enable one of ordinary skill in the art to practice the methods recited in the

claims without undue experimentation. In particular, the specification enables one of ordinary skill in the art, without requiring any undue experimentation, to release and amplify the recited immobilized probe, to label the resultant amplified probe, and to then hybridize the labeled probe to a target polynucleotide to immobilize the labeled probe and to detect the labeled probe.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful, false statements may jeopardize the validity of the application and any patent issuing therefrom.

Date: Nov 22, 2000

Mark A. Reynolds, Ph.D.

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10614–10619, October 1996
Biochemistry

Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes

(Human Genome Project/DNA chip/gene discovery/T cell)

MARK SCHENA*†, DARI SHALON‡, RENU HELLER*, ANDREW CHAI*, PATRICK O. BROWN§, AND RONALD W. DAVIS*

*Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305; ‡Synteni, Palo Alto, CA 94306; and §Department of Biochemistry and Howard Hughes Medical Institute, Beckman Center, Stanford University Medical Center, Stanford, CA 94305

Contributed by Ronald W. Davis, June 26, 1996

ABSTRACT **Microarrays containing 1046 human cDNAs of unknown sequence were printed on glass with high-speed robotics. These 1.0-cm² DNA "chips" were used to quantitatively monitor differential expression of the cognate human genes using a highly sensitive two-color hybridization assay. Array elements that displayed differential expression patterns under given experimental conditions were characterized by sequencing. The identification of known and novel heat shock and phorbol ester-regulated genes in human T cells demonstrates the sensitivity of the assay. Parallel gene analysis with microarrays provides a rapid and efficient method for large-scale human gene discovery.**

Biology has entered the genome era (1). Complete genome sequences for all of the model organisms and human will probably be available by the year 2003 (2). Torrents of human expressed sequence tags (ESTs) provide a starting point for elucidating the function of tens of thousands of cognate genes (3). Genome analysis will provide insights into growth, development, differentiation, homeostasis, aging, and the onset of diseases (1–3). A detailed understanding of the human genome will require the implementation of sophisticated methods for gene expression analysis and gene discovery.

Recently, a microarray-based method for high-throughput monitoring of plant gene expression was described (4). This "chip"-based approach involved using microarrays of cDNA clones as gene-specific hybridization targets to quantitatively measure expression of the corresponding plant genes (4, 5). A two-color fluorescence labeling and detection scheme facilitated sensitive differential expression analysis of different plant tissues (4, 5). The efficiency of this approach for studies in higher plants suggested the use of this method for human genome analysis (4–7). Here, we report the use of cDNA microarrays for human gene expression monitoring, biological investigation, and gene discovery.

MATERIALS AND METHODS

Human cDNA Clones. The cDNA library was made with mRNA from human peripheral blood lymphocytes transformed with the Epstein–Barr virus. Inserts >600 bp were cloned into the lambda vector λYES-R to generate 10^7–10^8 recombinants. Bacterial transformants were obtained by infecting *E. coli* strain JM107/λKC. Colonies were picked at random and propagated in a 96-well format, and minilysate DNA was prepared by alkaline lysis using REAL preps (Qiagen, Chatsworth, CA). Inserts were amplified by PCR in a 96-well format using primers (PAN132, 5′-CCTCTATACTTTAACGTCAAGG; and PAN133, 5′-TTGTGTGGAATTGTGAGCGG) complementary to the λYES polylinker and containing a six-carbon amino modification (Glen Research, Sterling, VA) on the 5′ end. PCR products were purified in a 96-well format using QIAquick columns (Qiagen).

Microarray Preparation. Amino-modified PCR products were suspended at a concentration of 0.5 mg/ml in 3× standard saline citrate (SSC) and arrayed from 96-well microtiter plates onto silylated microscope slides (CEL Associates, Houston) using high-speed robotics (4–7). A total of 1056 cDNAs, representing 1046 human clones and 10 *Arabidopsis* controls, were arrayed in 1.0-cm² areas. Printed arrays were incubated for 4 hr in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in H_2O for 1 min, and once for 5 min in sodium borohydride solution (1.0 g of NaBH4 dissolved in 300 ml of PBS and 100 ml of 100% ethanol). The arrays were submerged in H_2O for 2 min at 95°C, transferred quickly into 0.2% SDS for 1 min, rinsed twice in H_2O, air dried, and stored in the dark at 25°C.

Fluorescent Probes. Tissue mRNAs were purchased (CLONTECH). Jurkat mRNA was isolated as described by Schena *et al.* (4). Probes were made as described (4) with several modifications. The reverse transcriptase used here was Superscript II RNase H− (GIBCO). The Cy5-dCTP was purchased from Amersham. Each reverse transcription reaction contained 3.0 μg of total human mRNA. *Arabidopsis* control mRNAs were made by *in vitro* transcription of cloned HAT4, HAT22, and YesAt-23 cDNAs (4, 8, 9) using an RNA Transcription Kit (Stratagene). For quantitation, the mRNAs were doped into the reverse transcription reaction at ratios of 1:100,000, 1:10,000, and 1:1000 (wt/wt) respectively. Following the reverse transcription step, samples were treated with 2.5 μl of 1 M sodium hydroxide for 10 min at 37°C, then neutralized by adding 2.5 μl of 1 M Tris·HCl (pH 6.8) and 2.0 μl of 1 M HCl. Probe mixtures contained cDNA products derived from 3 μg of total mRNA, suspended in 5.0 μl of hybridization buffer (5× SSC plus 0.2% SDS).

Hybridization and Scanning. Probes were hybridized to 1.0-cm² microarrays under a 14 × 14 mm glass coverslip for 6–12 hr at 60°C in a custom-built hybridization chamber (4–7). Arrays were washed for 5 min at room temperature (25°C) in low stringency wash buffer (1× SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1× SSC/0.2% SDS). Arrays were scanned in 0.1× SSC using a fluorescence laser scanning device (4–7), fitted with a custom filter set (Chroma Technology, Brattleboro, VT). Accurate differential expression measurements (i.e., final fluorescence ratios) were obtained by taking the average of the ratios of two independent hybridizations.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EST, expressed sequence tag.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U56654–U56660).

†To whom reprint requests should be addressed. e-mail: schena@cmgm.stanford.edu.

Cell Culture. Jurkat cells were grown in a tissue culture incubator (37°C and 5% CO_2) in RPMI medium supplemented with 10% fetal bovine serum, 100 μg of streptomycin per ml, and 500 units of penicillin per ml. Heat shock corresponded to a 4-hr incubation at 43°C. Phorbol ester treated cells were grown for 4 hr in the presence of 50 ng of phorbol 12-myristate 13-acetate (PMA) per ml.

RNA Blotting. Dot blots were performed as described (4).

DNA Sequencing. Sequences were obtained using the PAN132 and PAN133 primers and a 373A automated sequencer, according to the instructions of the manufacturer (Applied Biosystems).

Computer Graphics and Informatics. Pseudocolor representations of fluorescent images were made with National Institutes of Health IMAGE software (version 1.52). Software for differential expression representations was purchased from Imaging Research (St. Catherine's, ON, Canada). Sequence searches were made to the nonredundant nucleotide data base at the National Center for Biotechnology Information (NCBI) using Macintosh BLAST software. The EST data base was accessed via the World Wide Web (http://www.ncbi.nlm.nih.gov/).

RESULTS

Gene Discovery and the Heat Shock Response. Microarrays were used to examine the heat shock response in cultured human T (Jurkat) cells. Control (37°C) and heat-treated (43°C) cells were harvested and lysed, and total mRNA from the two cell samples was labeled by reverse transcriptase incorporation of fluorescein- and Cy5-dCTP, respectively. In a second set of labeling reactions, the fluorescent groups were "swapped" such that samples from control and heat-treated samples were labeled with Cy5- and fluorescein-dCTP, respectively. Each pair of fluorescent probes was hybridized to a 1056-element microarray. The arrays were washed at high stringency and scanned with a confocal laser scanning device to detect emission of the two fluorescent groups.

Hybridization signals were observed to >95% of the human cDNA array elements, but not to any of the *Arabidopsis* negative controls (Fig. 1). Fluorescence intensities spanned more than three orders of magnitude for the 1046 array elements surveyed (Fig. 1). Comparative expression analysis of heat shocked versus control cells in the two experiments revealed 17 array elements that displayed altered fluorescence ratios of ≥2.0-fold (Figs. 1 and 2*A*). Of the 17 putative differentially expressed genes, 11 were induced by heat shock treatment and 6 displayed modest repression (Figs. 1 and 2*A*).

To determine the identity of the heat-regulated genes, cDNAs corresponding to each of the 17 array elements were sequenced on the proximal and distal end. Data base searches revealed perfect matches for 14 of the 17 clones, and in each case proximal and distal cDNA sequences mapped to the same gene (Table 1). Of the 1046 human genes examined on the microarray, the five most highly induced in heat-treated cells were heat shock protein 90α (hsp90α), dnaJ, hsp90β, polyubiquitin, and t-complex polypeptide-1 (tcp-1) (Table 1). Three of the 17 clones did not match any entry in the public data base, though one of the clones (B7) exhibited significant homology to an EST from *Caenorhabditis elegans* (Table 1). Each of the novel sequences (B7-B9) exhibited ≈2-fold induction (Table 1) and relatively low-level expression (Table 2).

To confirm the microarray results, mRNA levels for each of the genes were measured by RNA blotting. Each of the genes that displayed heat shock induction, including the three novel

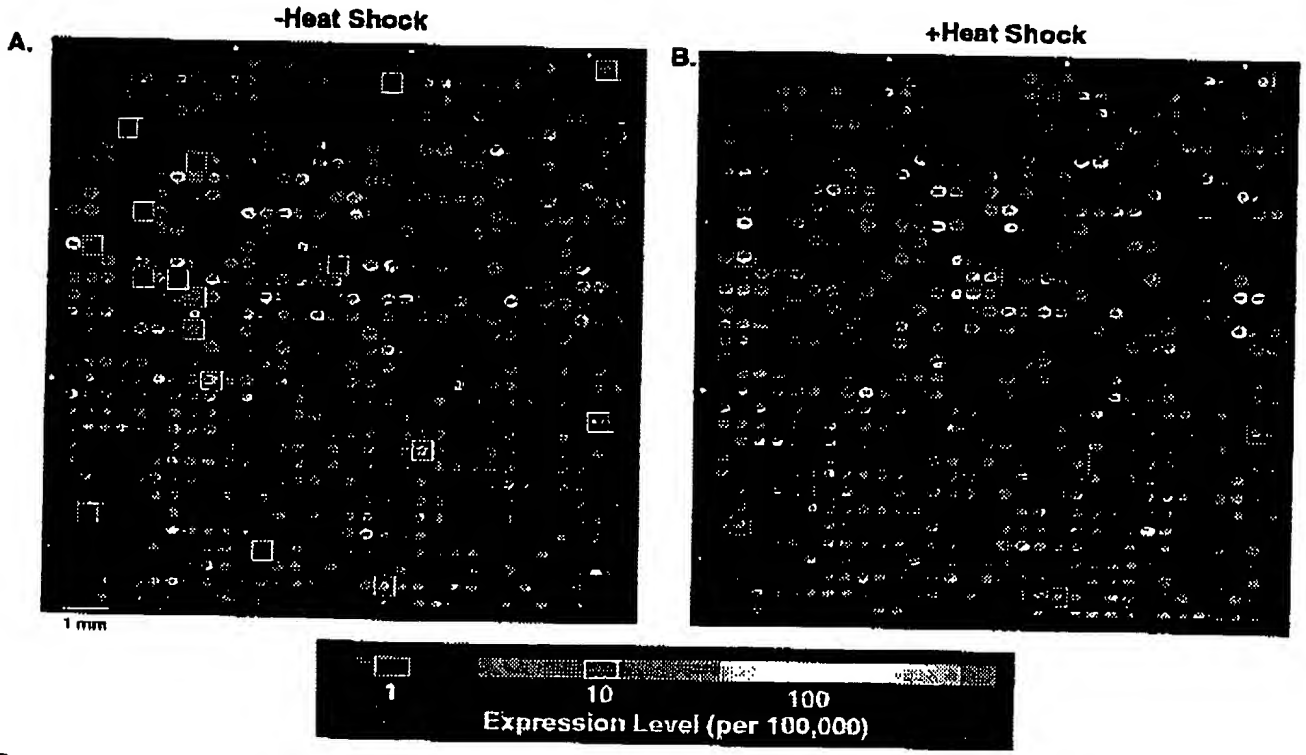

FIG. 1. Human gene expression monitored on a microarray. Fluorescent scans represented in a pseudocolor scale correspond to expression levels. The array contains 10 *Arabidopsis* controls (upper left corner, elements 1–10) and 1046 human peripheral blood cDNAs. Fluorescent probes were prepared by labeling mRNA from Jurkat cells grown at 37°C (−Heat Shock, *A*) or 43°C (+Heat Shock, *B*). Array elements that display altered fluorescence intensity (white boxes) corresponded to genes activated (red boxes) or repressed (green boxes) by heat shock. The color bar was calibrated in separate experiments using known quantities (wt/wt) of *Arabidopsis* control mRNAs added to the labeling reaction. Microarray rows (at left) and columns (at the top) are demarcated at 10 element increments (white circles). (Bar = 1 mm.)

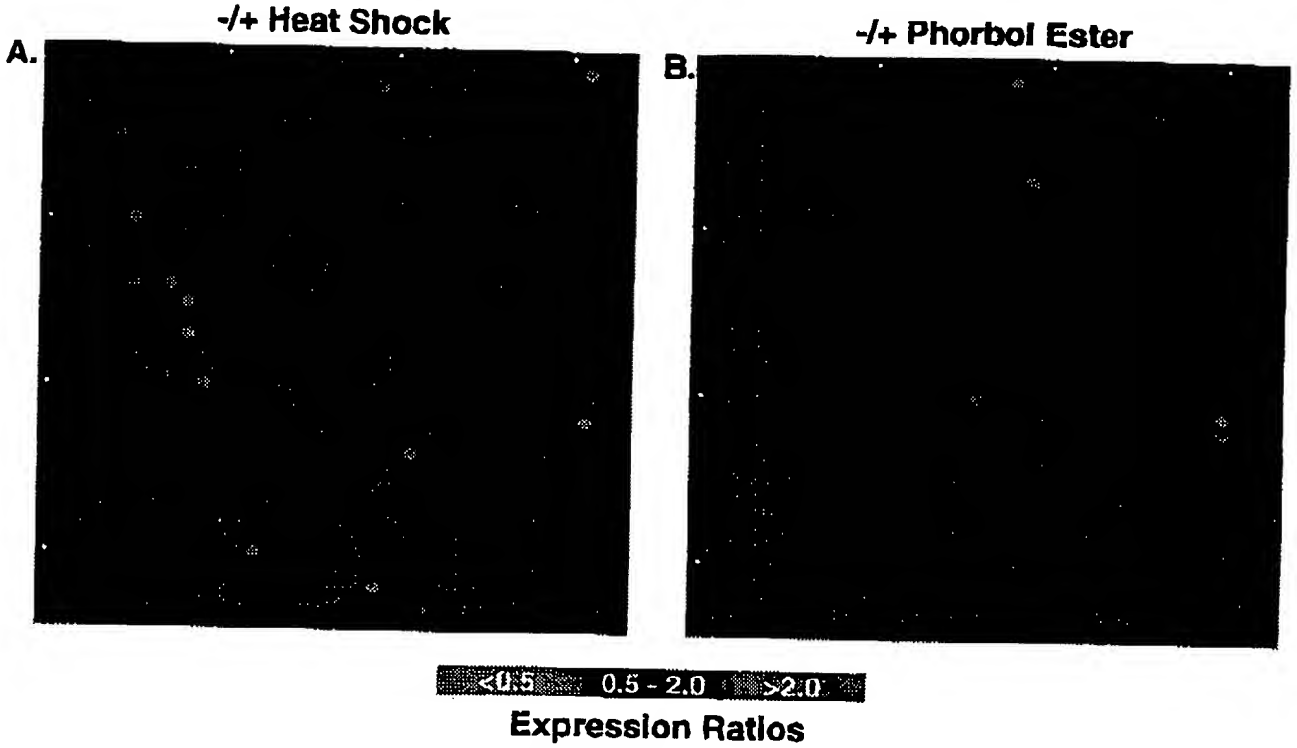

FIG. 2. Elemental displays of activated and repressed genes. Fluorescence ratios of two-color microarray scans (Fig. 1) are depicted schematically. Fluorescein-labeled probes from Jurkat cells subjected to (*A*) heat shock or (*B*) phorbol ester treatment were compared with Cy5-labeled probes from untreated cells. In a second set of reactions, the fluorescent groups were swapped (see text). The data represent the average of the ratios from two hybridizations, excluding values in which the difference of the two ratios was greater than half the average ratio. The color bar corresponds to expression ratios, which are independent of the absolute expression level of a given gene.

Table 1. Microarray elements corresponding to differentially expressed genes

Clone	Row	Column	Ratio	Blast identity	Accession no.
B1	24	21	0.5	CYC oxidase III	J01415, J01415
B2	1	31	0.5	β-Actin	NR, X00351
B3	15	8	0.5	CYC oxidase III	J01415, J01415
B4	32	19	0.5	CYC oxidase III	J01415, J01415
B5	17	8	0.5	CYC oxidase III	J01415, J01415
B6	22	31	0.5	β-Actin	NR, X00351
B7*	5	4	2.0	Novel†	U56653, U56654
B8	2	19	2.0	Novel†	U56655, U56656
B9	14	5	2.2	Novel†	U56657, U56658
B10	7	8	2.4	Polyubiquitin	X04803, X04803
B11	12	2	2.4	TCP-1	X52882, X52882
B12	28	2	2.5	Polyubiquitin	M17597, M17597
B13	14	7	2.5	Polyubiquitin	X04803, X04803
B14	20	9	2.6	HSP90β	M16660, M16660
B15	30	12	4.0	DnaJ homolog	D13388, D13388
B16	10	5	5.8	HSP90α	X07270, X07270
B17	13	16	6.3	HSP90α	M27024, X15183
B18	7	19	2.0	β_2-microglobulin	S54761, M30683
B19	21	30	2.1	Novel†	U56659, U56660
B20	3	26	2.2	β_2-microglobulin	S54761, M30683
B21	1	18	2.6	PGK	M11968, L00160
B22	22	30	3.5	NF-κB1	Z47744, M55643
B23	20	16	19	PAC-1	L11329, L11329

Clone name, array position (Fig. 1), fluorescence ratio, sequence identity, and acession number of cDNAs that manifested a differential expression pattern with probes prepared from heat shock- (B1-17) or phorbol ester-treated (B18-23) Jurkat cells. Clones showing >98% identity over 300 nucleotides were assumed to be identical to known sequences. All genes are nuclear except CYC oxidase III (mitochondrial). Accession numbers reflect the highest score for proximal and distal sequence traces, respectively. CYC, cytochrome *c*; TCP-1, T-complex polypeptide; HSP, heat shock protein; PGK, phosphoglycerate kinase; NF-κB, nuclear factor-kappaB; PAC-1, phosphatase of activated cells; and NR, trace not readable due to the presence of poly(A)+ tract.

*B7 is 67% identical to an EST from *C. elegans* (D76025).

†No match in the public data bases.

Table 2. Human gene expression monitored by microarray and RNA blot analyses

Clone	Blast identity	Expression level, per 10^5 mRNAs			
		Microarray	Ratio	RNA blot	Ratio
B1	CYC oxidase III	92/46	0.5	100/80	0.8
B2	β-Actin	240/120	0.5	270/280	1.0
B3	CYC oxidase III	36/18	0.5	ND	ND
B4	CYC oxidase III	76/38	0.5	ND	ND
B5	CYC oxidase III	62/31	0.5	ND	ND
B6	β-Actin	180/89	0.5	ND	ND
B7	Novel (weakly to D76026)	1.3/2.6	2.0	0.77/1.8	2.3
B8	Novel	2.0/4.0	2.0	1.5/3.4	2.3
B9	Novel	0.8/1.8	2.2	1.2/1.8	1.5
B10	Polyubiquitin	0.8/1.9	2.4	25/89	3.6
B11	TCP-1	2.3/5.5	2.4	7.1/27	3.8
B12	Polyubiquitin	0.8/2.0	2.5	ND	ND
B13	Polyubiquitin	1.7/4.3	2.5	ND	ND
B14	HSP90β	75/200	2.6	30/120	4.0
B15	DnaJ homolog	1.0/4.0	4.0	1.6/13	8.1
B16	HSP90α	0.6/3.5	5.8	3.2/29	9.1
B17	HSP90α	0.8/5.0	6.3	8.6/62	7.2
B18	β_2-microglobulin	1.0/2.0	2.0	5.4/15	2.8
B19	Novel	1.2/2.5	2.1	4.5/9.5	2.5
B20	β_2-microglobulin	2.7/5.9	2.2	ND	ND
B21	Phosphoglycerate kinase	2.4/6.2	2.6	4.7/9.2	2.0
B22	NF-KB1	1.7/6.0	3.5	0.65/4.7	7.2
B23	PAC-1	0.5/9.5	19	0.21/15	71

Shown are expression levels per 100,000 mRNAs (wt/wt) of genes assayed with a microarray (Fig. 1) or RNA blot. Ratios correspond to values from cells subjected to heat shock (B1-17) or phorbol ester treatment (B18-23) relative to untreated cells. Clone and gene names are given in Table 1. ND, not determined.

sequences, exhibited elevated mRNA levels by dot blot analysis (Table 2). In all cases, expression ratios as determined by the two procedures differed by <2-fold for the genes identified in the heat shock experiments (Table 2). The two assays differed more widely in terms of assessing absolute expression levels; nonetheless, absolute expression as monitored on a microarray typically correlated with RNA blots to within a factor of five (Table 2).

Phorbol Ester Signaling. To explore a signaling pathway distinct from the heat shock response, microarrays were used to examine the cellular effects of phorbol ester treatment. Jurkat cells were treated with phorbol ester, harvested, lysed, and used as a source of mRNA. Samples of mRNA from untreated or phorbol ester-stimulated cells were labeled with reverse transcriptase. The probes were mixed, hybridized to microarrays, and scanned for fluorescence emission of the two fluorescent groups. A total of six array elements displayed ≥2.0-fold elevated signals with probes from phorbol ester-treated cells relative to control samples (Fig. 2*B*).

To determine the identity of the phorbol ester-induced genes, clones corresponding to the six array elements were sequenced. Data base searches revealed perfect matches for five of the six sequences (Table 1). The two most highly induced genes were the *PAC-1* tyrosine phosphatase and nuclear factor-kappa B1 (*NF-κB1*); modest activation was observed for phosphoglycerate kinase and β_2-microglobulin (Table 1). One remaining clone (B19) did not match any entry in the public data base (Table 1). B19 displayed a 2.1-fold induction and, similar to the novel heat shock genes, a relatively low absolute expression level (Tables 1 and 2). All six of the phorbol ester-inducible genes displayed increased steady-state mRNA levels by RNA blotting (Table 2). *PAC-1* expression (Fig. 1; Table 2) defined a detection limit of ≈1:500,000 for the assay.

Transcript Imaging in Human Tissues. To determine whether microarrays could be used to monitor expression in human tissues, probes were prepared from human bone marrow, brain, prostate, and heart by labeling each mRNA sample with Cy5-dCTP. In a separate reaction, a control probe was prepared by labeling Jurkat mRNA with fluorescein-dCTP. The four Cy5-labeled probes were each mixed with an aliquot of the fluorescein-labeled control sample, and the four mixtures were hybridized to separate microarrays. The arrays were washed and scanned for fluorescence emission, and hybridization signals for each of the tissues samples were normalized to the Jurkat control to generate an expression profile for each of the 1046 clones present on the array.

Detectable expression was observed for all 15 of the heat shock and phorbol ester-regulated genes in the four tissue types examined (Fig. 3). In general, the expression level of each gene in Jurkat cells correlated rather closely with expression in the four tissues (Table 2; Fig. 3). Genes encoding β-actin and cytochrome *c* oxidase, the two most highly expressed of the 15 genes in Jurkat cells (Table 2), were highly expressed in bone marrow, brain, prostate, and heart (Fig. 3*A*). Expression of cytochrome *c* oxidase, hsp90α, and the novel B7 sequence was significantly greater in heart than in the other tissues (Fig. 3).

DISCUSSION

Many of the heat shock genes identified in this study encode factors that function either as molecular "chaperones" (HSP90α, HSP90β, DnaJ, TCP-1) or as mediators of protein degradation (polyubiquitin). The identification of these sequences is consistent with the biochemical basis of heat shock induction (10–15). Proteins undergo denaturation at elevated temperatures, and those that fail to maintain proper conformation must be selectively degraded (10–15). It will be interesting to determine whether the three novel heat shock-inducible sequences (B7–B9) mediate protein folding and turnover or possess some other biochemical activity. Complete nucleotide sequence determination, conceptual translation, expression monitoring, and biochemical analysis should provide a detailed functional understanding of these genes.

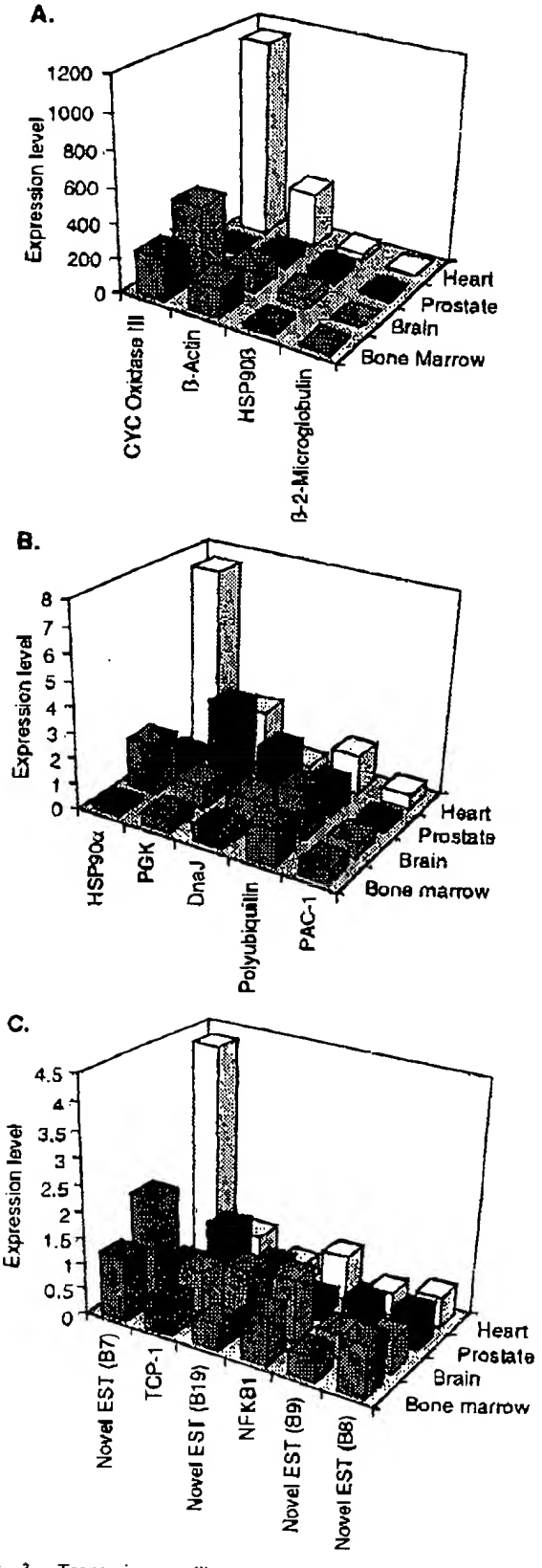

FIG. 3. Transcript profiles of heat shock and phorbol ester-regulated genes. Gene expression levels per 100,000 mRNAs (z-axes) are shown for 15 genes (Table 1) in human bone marrow (red), brain (green), prostate (blue), and heart (yellow). Genes are grouped according to expression levels (*A–C*).

Phorbol ester, a potent activator of protein kinase C (16, 17), induced a set of genes distinct from those involved in the heat shock pathway. The most highly induced gene identified in this study, *PAC-1*, encodes a nuclear tyrosine kinase that may play a role in regulating transcription and cell cycle progression (18). NF-κB1, a second phorbol ester-inducible gene, is an intensively studied member of the Rel transcription factor family (19–21). The Rel proteins are activated by a large number of stimuli, including phorbol esters, cytokines, bacterial and viral pathogens, and ultraviolet light (19–21). Modest activation was observed for three sequences not known to be inducible by phorbol esters, including phosphoglycerate kinase, β_2-microglobulin, and a novel human gene (B19). Extensive expression monitoring with microarrays should assist in understanding how each of these genes integrate into the highly complex phorbol ester signaling pathway.

It is striking that four novel human genes were discovered with an array of 1000 randomly chosen clones, particularly because the heat shock and phorbol ester signaling pathways have been so intensively studied (10–21). The facile discovery of these sequences underscores the fact that microarrays can be used for gene discovery in the absence of any sequence information. By this approach, clones are chosen at random from any library of interest and only those clones that display interesting expression patterns are sequenced and characterized. This parallel assay, coupled with a modest DNA sequencing facility, allows high-throughput human genome expression analysis and gene discovery.

Genes that are activated or repressed by a given stimulus provide functional clues to the cellular pathway involved (22–24). Detailed examination of these gene expression "signatures" can provide a dynamic view of the mode of action of a given signaling substance (22–24). Microarrays may thus allow rapid mechanistic examination of hormones, drugs, elicitors, and other small molecules; moreover, functional analysis of transcription factors, kinases, growth factors, cytokines, receptors, and other gene products should be possible. Efforts are underway to develop mRNA amplification strategies to enable probe preparation from minute tissue samples. This capability might allow for high-throughput patient screening in a clinical setting.

The current detection limit of the assay allows monitoring of transcripts that represent ≈1:500,000 (wt/wt) of the total mRNA. This 10-fold increase in sensitivity compared with the original report (4) was achieved largely by modifying the coupling chemistry, which reduced background fluorescence. The significance of this improvement is considerable in that approximately half the human genes identified in this study, including all four novel sequences, exhibited expression levels below the original detection limit of 1:50,000 (4).

The ability to detect 2-fold changes in expression was achieved by the use of two-color fluorescence in the labeling and detection schemes, digitized data collection, and custom software. The importance of this capability is underscored by the fact that nearly all of the genes examined here exhibited <6-fold changes in expression. The four novel genes, which showed ≤2.2-fold activation, were probably overlooked in previous screens that used conventional differential expression techniques. It may be possible to further improve the precision of the microarray assay by the use of closely related fluorescent analogs, such as Cy3 and Cy5, in the labeling and hybridization reactions.

Microarrays offer a number of advantages over other potential high-capacity approaches to expression analysis. The chip-based approach enables small hybridization volumes, high array densities, and the use of fluorescence labeling and detection schemes. These features provide a set of performance specifications that are unattainable with filter-based approaches (25, 26). The use of cDNA clones provides hybridization specificity that is not readily attained with oligo-

nucleotide arrays (27–30). The parallel format of the assay provides a simultaneous differential expression readout for >1000 genes. This contrasts with sequencing-based methods, which require serial data collection for expression analysis (31, 32). A commercial source of cDNA microarrays would greatly speed the use of a chip-based approach to expression analysis.

The availability of large numbers of ESTs (3) provides a rich resource of human cDNA clones for microarraying. The >400,000 ESTs in the public data bases represent a significant subset of all human genes (3, 33). Microarrays of thousands of ESTs will provide a powerful analytical tool for future human gene expression studies. The ≈100,000 genes in the human genome (2, 33) emphasize the need for microarrays of greater density. Attempts to improve microdeposition techniques are underway and should allow construction of arrays containing a complete set of human gene targets (http://cmgm.stanford.edu/~schena/). Microarrays of ≈100,000 cDNA elements would allow expression monitoring of the entire human genome in a single hybridization. This capacity, coupled with detailed biochemical analysis of the individual gene products, would greatly speed the functional analysis of the human genome.

We thank S. Elledge (selledge@bcm.tmc.edu) for the human cDNA library, Qiagen representatives for help with plasmid purification, and A. J. Smith and colleagues at the Protein and Nucleic Acid (PAN) facility (Stanford) for oligonucleotide synthesis and DNA sequencing. We also thank members of the Davis, Brown, and Smith laboratories for critical comments and helpful discussions and Synteni employees for technical assistance. Support for R.W.D. was provided by the National Science Foundation (MCB9106011) and National Institutes of Health (R37HG00198) and for P.O.B. by the National Institutes of Health (3R21HG00450) and Howard Hughes Medical Institute. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

1. Watson, J. D. (1993) *Gene* **135**, 309–315.
2. Collins, F. S. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 10821–10823.
3. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R. & Venter, J. C. (1991) *Science* **252**, 1651–1656.
4. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* **270**, 467–470.
5. Shalon, D. (1996) Ph.D. thesis (Stanford University).
6. Schena, M. (1996) *BioEssays* **18**, 427–431.
7. Shalon, D., Smith, S. J. & Brown, P. O. (1996) *Genome Res.* **6**, 639–645.
8. Schena, M. & Davis, R. W. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 8393–8397.
9. Schena, M. & Davis, R. W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3894–3898.
10. Jindal, S. (1996) *Trends Biotechnol.* **14**, 17–20.
11. Wilkinson, K. D. (1995) *Annu. Rev. Nutr.* **15**, 161–189.
12. Jakob, U. & Buchner, J. (1994) *Trends Biochem. Sci.* **19**, 205–211.
13. Becker, J. & Craig, E. A. (1994) *Eur. J. Biochem.* **219**, 11–23.
14. Cyr, D. M., Langer, T. & Douglas, M. G. (1994) *Trends Biochem. Sci.* **19**, 176–181.
15. Craig, E. A., Weissman, J. S. & Horwich, A. L. (1994) *Cell* **78**, 365–372.
16. Newton, A. C. (1995) *J. Biol. Chem.* **270**, 28495–28498.
17. Nishizuka, Y. (1995) *FASEB J.* **9**, 484–496.
18. Rohan, P. J., Davis, P., Moskaluk, C. A., Kearns, M., Krutzsch, H., Siebenlist, U. & Kelly, K. (1993) *Science* **259**, 1763–1766.
19. Thanos, D. & Maniatis, T. (1995) *Cell* **80**, 529–532.
20. Baeuerle, P. A. & Henkel, T. (1994) *Annu. Rev. Immunol.* **12**, 141–179.
21. Liou, H.-C. & Baltimore, D. (1993) *Curr. Opin. Cell Biol.* **5**, 477–487.
22. Cohen, G. B., Ren, R. & Baltimore, D. (1995) *Cell* **80**, 237–248.
23. Chan, A. C., Desai, D. M. & Weiss, A. (1994) *Annu. Rev. Immunol.* **12**, 555–592.
24. Crabtree, G. R. & Clipstone, N. A. (1994) *Annu. Rev. Biochem.* **63**, 1045–1083.
25. Gress, T. M., Hoheisel, J. D., Lennon, G. G., Zehetner, G. & Lehrach, H. (1992) *Mamm. Genome* **3**, 609–619.
26. Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A.-M., Jordan, B. R. & Nguyen, C. (1996) *Nucleic Acids Res.* **24**, 1435–1442.
27. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991) *Science* **251**, 767–773.
28. Southern, E. M., Maskos, U. & Elder, J. K. (1992) *Genomics* **13**, 1008–1017.
29. Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994) *Nucleic Acids Res.* **22**, 5456–5465.
30. Matson, R. S., Rampal, J., Pentoney, S. L., Jr., Anderson, P. D. & Coassin, P. (1995) *Anal. Biochem.* **224**, 110–116.
31. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* **270**, 484–487.
32. Adams, M. D. (1996) *BioEssays* **18**, 261–262.
33. Fields, C., Adams, M. D., White, O. & Venter, J. C. (1994) *Nat. Genet.* **7**, 345–346.

Expression monitoring by hybridization to high-density oligonucleotide arrays

:vid J. Lockhart*, Helin Dong, Michael C. Byrne[1], Maximillian T. Follettie[1], Michael V. Gallo[1], Mark S. Chee, Michael Mittmann, Chunwei Wang, Michiko Kobayashi[1], Heidi Horton[1], and Eugene L. Brown[1]

*Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051. [1]Genetics Institute, 87 CambridgePark Dr., Cambridge, MA 02140. *Corresponding author (e-mail: david_lockhart@affymetrix.com).*

Received 21 August 1996; accepted 7 October 1996

The human genome encodes approximately 100,000 different genes, and at least partial sequence information for nearly all will be available soon. Sequence information alone, however, is insufficient for a full understanding of gene function, expression, regulation, and splice-site variation. Because cellular processes are governed by the repertoire of expressed genes, and the levels and timing of expression, it is important to have experimental tools for the direct monitoring of large numbers of mRNAs in parallel. We have developed an approach that is based on hybridization to small, high-density arrays containing tens of thousands of synthetic oligonucleotides. The arrays are designed based on sequence information alone and are synthesized in situ using a combination of photolithography and oligonucleotide chemistry. RNAs present at a frequency of 1:300,000 are unambiguously detected, and detection is quantitative over more than three orders of magnitude. This approach provides a way to use directly the growing body of sequence information for highly parallel experimental investigations. Because of the combinatorial nature of the chemistry and the ability to synthesize small arrays containing hundreds of thousands of specifically chosen oligonucleotides, the method is readily scalable to the simultaneous monitoring of tens of thousands of genes.

Keywords: gene expression, mRNA abundance, parallel, quantitative, hybridization, combinatorial synthesis, oligonucleotide arrays, DNA chips

Only a subset of all encoded genes is expressed in any given cell, and the levels and timing of expression govern cellular development, differentiation, function, and physiology. We describe a method for the simultaneous monitoring of the expression levels of many genes in parallel. The approach involves the determination of the relative concentrations of mRNAs based on hybridization of entire mRNA populations to high-density arrays of synthetic oligonucleotides. The arrays used contain either 16,000 or more than 65,000 different 20-mer oligonucleotides of defined sequence in an area of 1.6 cm^2. The surface-bound oligonucleotides are referred to as probes because they serve to probe, or interrogate, the composition of the RNA population. The DNA probes are synthesized directly on derivatized glass slides using a combination of photolithography and oligonucleotide chemistry[1-3], and each of the probe sequences can be independently specified. The arrays are made in large numbers in a controlled environment and are highly reproducible.

Various methods are available for detecting or quantitating mRNAs. Methods such as Northern blots and nuclease protection have the disadvantage of being inherently serial, involving measuring a single mRNA at a time, and of being difficult to automate. Differential display[4] of amplified subsets of RNAs on a sequencing gel allows a broad search for expression differences, but the results generally are not quantitative, false positives are common, and characterization of positives requires additional cloning and sequencing. Sequencing of cDNA libraries[5,6] is a more direct approach, but requires a great deal of sequencing and is not sensitive to the presence of less abundant messages. The Serial Analysis of Gene Expression (SAGE) method of Velculescu et al.[7] is a clever and efficient variation of the cDNA sequencing approach. The method involves fairly complicated sample preparation procedures, still requires a large amount of sequencing, and tends to be laborious and not particularly sensitive.

Array-based methods that involve the spotting of multiple clones or cDNAs onto nylon membranes[10-17] or modified glass microscope slides[18,19] have been described. These approaches have the inherent advantages of being parallel with a direct and rapid readout of the hybridization results. The disadvantages are that in order to monitor many genes, a large number of cDNAs or PCR products must be prepared, purified, quantitated, catalogued, and spotted onto a solid support. If the cDNAs are derived from a cDNA library, low abundance cDNAs are unlikely to be spotted and the library must be normalized to reduce the redundant spotting of cDNAs from highly expressed genes. One potential advantage of using spotted cDNAs is that they might be expected to hybridize with greater affinity and specificity than shorter oligonucleotide probes. However, at present this does not appear to be the case. In the method of Schena et al.[18] and Shalon et al.[19], this may be because the spotted cDNAs are attached to the surface of polylysine-coated glass slides through a large number of interactions between the negatively charged phosphate backbone and the positively charged surface-bound lysine sidechains. These interactions tend to reduce the conformational freedom of the bound cDNAs and hence their affinity for complementary molecules in solution.

Our approach differs from spotting methods in that mRNA

The development of VLSIPS™ technology has provided methods for making very large arrays of oligonucleotide prob in very small arrays. See U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092, each of which is incorporated herein by reference. U.S. Patent application Serial No. 082,937, filed June 25, 1993, describ methods for making arrays of oligonucleotide probes that can be used to provide the complete sequence of a target nucleic acid and to detect the presence of a nucleic acid containing specific nucleotide sequence.

Microfabricated arrays of large numbers of oligonucleotide probes, called "DNA chips" offer great promi for a wide variety of applications. New methods and reagents are required to realize this promise, and the present invention helps meet that need.

SUMMARY OF THE INVENTION

The invention provides several strategies employing immobilized arrays of probes for comparing a reference sequence of known sequence with a target sequence showing substantial similarity with the reference sequence, but differing in the presence of, *e.g.*, mutations. In a first embodiment, the invention provides a tiling strategy employi an array of immobilized oligonucleotide probes comprising at least two sets of probes. A first probe set comprises a plurality of probes, each probe comprising a segment of at least three nucleotides exactly complementary to a subsequen of the reference sequence, the segment including at least on interrogation position complementary to a corresponding nucleotide in the reference sequence. A second probe set comprises a corresponding probe for each probe in the first probe set, the corresponding probe in the second probe set being identical to a sequence comprising the corresponding probe from the first probe set or a subsequence of at least three nucleotides thereof that includes the at least one interrogation position, except that the at least one